DOCUMENT RESUME

ED 264 264 TM 850 749

AUTHOR McDonald, Roderick P.
TITLE Unidimensional and Multidimensional Models for Item Response Theory.
PUB DATE Jul 82
NOTE 23p.; In: Item Response Theory and Computerized Adaptive Testing Conference Proceedings (Wayzata, MN, July 27-30, 1982) (TM 850 744).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Factor Analysis; Higher Education; Item Analysis; *Latent Trait Theory; *Mathematical Models; Postsecondary Education; *Psychometrics; *Regression (Statistics); Research Problems; Sample Size
IDENTIFIERS Law School Admission Test; *Unidimensional Scaling

ABSTRACT

This paper provides an up-to-date review of the relationship between item response theory (IRT) and (nonlinear) common factor theory and draws out of this relationship some implications for current and future research in IRT. Nonlinear common factor analysis yields a natural embodiment of the weak principle of local independence in appropriate loss functions that can be used to fit item response models and to assess their adequacy as descriptions of the data. The recognition of latent traits as synonymous with common factors gives appropriate guidance for interpretation in both unidimensional and multidimensional versions. Fixed-regressors theory based on the McDonald's (1979) treatment of common factor analysis was viewed, in principle, as applicable to any prescribed unidimensional model in item response theory, though so far it has been applied only to the rather inappropriate polynomial model. The advantages of this approach appear to be that it is not limited to the use of a linear combination rule for multiple latent traits and that it easily permits the introduction of interaction terms. Applications of the fixed-regressors treatment are subject to limitations of sample size, which may prove severe in the case of binary data, where large sample sizes are generally desirable. (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED264264

Unidimensional and Multidimensional Models for Item Response Theory

Roderick P. McDonald
Macquarie University

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☒ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

McDonald (1967a, 1967b) described a very general model for nonlinear common factor analysis and showed that its special cases include the major models in latent trait, or item response, theory. Also, Lord and Novick's (1968) widely influential account of item response theory (IRT) sets out a number of relationships betweeen latent trait theory and (linear or nonlinear) common factor theory. Nevertheless, it seems that the substantial mathematical overlap between IRT and common factor theory is still not widely appreciated. The object of this paper is to provide an up-to-date review of the relationship between IRT and (nonlinear) common factor theory and to draw out of this relationship some implications for current and future research in IRT.

At the outset, there seems to be a notable difference between latent trait theory and common factor theory. Latent trait theory is defined by the principle of local independence, while common factor theory is based upon a weaker principle. In the context adopted in this paper, that of test theory, the principle of local independence states that conditional upon any fixed value of a (scalar or k-component vector) latent trait $\underset{\sim}{\theta}$ associated with the members of a population of examinees, the scores of the examinees on a set of n items or tests are mutually statistically independent. The principle actually constitutes the definition of the latent trait(s) $\underset{\sim}{\theta}$. That is, $\underset{\sim}{\theta}$ is a vector of latent traits if and only if it is a set of quantities associated with the examinees such that the responses are mutually statistically independent for any fixed value of $\underset{\sim}{\theta}$. The English language labels "latent," "trait," and commonly associated words such as "underlying," "unobservable," and "ability" do not contain or convey any part of this definition and do not add anything to it. Since the latent traits explain all associations between the item responses, a latent trait is ordinarily interpreted as a measure of a trait in the psychological sense of the word, that is, an abstract attribute of the examinees which the items measure in common. In particular, in the case of a cognitive test the trait is commonly interpreted as an "ability."

Common factor theory, together with its common factor scores, is defined by the principle that conditional upon any value of a k-component vector $\underset{\sim}{\theta}$ of common factor scores, the scores of examinees on a set of items or tests are mutually uncorrelated. McDonald (1979, 1981) has suggested that the principle of local independence is described as taking two forms; in the strong form the responses are conditionally independent, and in the weak form they are conditionally uncorrelated. It is reasonable to assert that in all work with latent traits or common factors the strong form of the principle is assumed, although

TM 850 749

ordinarily the implied weak form, at most, is tested verifying that the residual covariances of the items are acceptably small but not checking whether the higher moments of their joint distribution are, conditionally, products of their univariate moments. Unfortunately, in a great deal of work on IRT (and occasionally in common factor analysis) the investigators do not even obtain the residual covariances of the items, or any other statistic such as a function of the residual covariances which measures the extent to which the defining principle of the model is satisfied. Since it is always easy to compute the residual covariances, it is difficult to understand why this is not standard practice. Possibly a failure to examine residual covariances in IRT models follows from a failure to recognize them as common factor models. If so, it can immediately be suggested that a central implication of the factor analytic perspective on IRT is that it should always prove informative to compute and examine the residual covariance matrix of the items when fitting such "standard" models as the Rasch model, the normal ogive model, and the logistic model, as well as other models to be discussed below.

The general nonlinear common factor model (McDonald, 1962a, 1967a, 1967b) may be expressed as the nonlinear multivariate regression equation

$$\hat{\underset{\sim}{y}} = \mathcal{E}\{\underset{\sim}{y}|\underset{\sim}{\theta}\} = \underset{\sim}{\phi}(\underset{\sim}{\theta}) \qquad [1]$$

$$\hat{y}_j = \mathcal{E}\{y_j|\theta\} = \phi_j(\theta),\ j = 1, \ldots, n, \qquad [2]$$

where

$\underset{\sim}{y}' = [y_1, \ldots, y_n]$ is a vector of test or item scores;

$\underset{\sim}{\phi}$ is a vector of functions, in general, nonlinear; and

$\underset{\sim}{\theta}$ is a $k \times 1$ vector of latent traits or common factors defined by the principle of local independence, at least in its weak form.

A vector of residuals $\underset{\sim}{e}$ is defined by

$$\underset{\sim}{e} = \underset{\sim}{y} - \mathcal{E}\{\underset{\sim}{y}|\underset{\sim}{\theta}\} \qquad [3]$$

or

$$e_j = y_j - \mathcal{E}\{y_j|\underset{\sim}{\theta}\},\ j = 1, \ldots, n\ . \qquad [4]$$

It then follows without further assumptions that the covariance structure of the observed scores is given by

$$\underset{\sim}{C} = \mathrm{Cov}\{\underset{\sim}{y}\} = \mathrm{Cov}\{\underset{\sim}{\phi}\} + \underset{\sim}{U}^2\ , \qquad [5]$$

where $\underset{\sim}{U}^2$, diagonal, nonnegative definite, is the residual covariance matrix and contains unique variances in the usual sense of the term in common factor theory (see McDonald, 1967b). The residual is ordinarily composed of a stable specific part and an unstable error of replication. It is important to recognize that, in general, the residual variances are not required to be homoscedastic and that binary items, as well as "quantitative" test scores, have both a specific and an error component in the sense understood from common factor theory.

McDonald (1982b, in press) suggests a fundamental classification of the family of all latent trait or common factor models into (1) strictly linear models, in which the functions $\phi(\underset{\sim}{\theta})$ are linear both in the coefficients of the regressions on the latent traits and in the latent traits themselves; (2) wide-sense linear models, in which the functions $\phi(\underset{\sim}{\theta})$ are linear in the coefficients of the regressions but not in the latent traits; and (3) strictly nonlinear models, which, loosely speaking, are nonlinear in both. It is safer to define a strictly nonlinear model as one that cannot be expressed as a wide-sense linear model with a finite number of terms, since any strictly nonlinear model may in principle be approximated as closely as desired by a wide-sense linear model using harmonic analysis (McDonald, 1967a).

The linear common factor model

$$\hat{y}_j = \sum_{p=1}^{k} f_{jp}\theta_p \qquad [6]$$

is the obvious example of a strictly linear model. Lazarsfeld's (Lazarsfeld & Henry, 1968) latent class model is also strictly linear.

The exploratory methods for nonlinear factor analysis described by McDonald (1962b, 1967b, 1967c, 1967d) essentially consist of wide-sense linear models, in which the regression function is a linear combination of polynomial functions, and, possibly, products of latent traits. A central result of the theory is the demonstration (McDonald, 1967b) that if a wide-sense linear model contains $\underline{r}$ functions of the $\underline{k}$ latent traits, then the covariance structure of the observed variables can be expressed as

$$\underset{\sim}{C} = \text{Cov}\{\underset{\sim}{y}\} = \underset{\sim}{B}\underset{\sim}{B}' + \underset{\sim}{U}^2 \qquad [7]$$

where $\underset{\sim}{B}$ is an $n \times r$ matrix. Consequently, the model cannot be distinguished from the strictly linear model with $\underline{r}$ common factors on the basis of the covariance structure alone. This fact serves to explain the problem of difficulty factors in the factor analysis of binary items (McDonald, 1967a; McDonald & Ahlawat, 1974).

The best-known examples of strictly nonlinear latent trait models would be the almost interchangeable normal ogive and logistic models of IRT. Other less well-known examples are the latent distance models and latent content models (Lazarsfeld & Henry, 1968). Neither of these, however, seems of interest in the context of test theory.

In principle, any latent trait model can be treated as a random-regressors model or a fixed-regressors model, the choice being dictated by a number of mathematical considerations. The least of these, curiously enough, is the fundamental logic of the distinction. In the random regressors model, the $\underline{n}$ observed variables $y_1, \ldots, y_n$ and the $\underline{k}$ latent traits $\theta_1, \ldots, \theta_k$, are assumed to constitute $n + k$ random variables, with a joint distribution in $n + k$ dimensions. A sample of size N drawn from this distribution consists of N random vectors $\underset{\sim}{y}_1, \ldots, \underset{\sim}{y}_N$ independently and identically distributed like $\underset{\sim}{y}$, and there

is no question of estimating the unknown corresponding $\underset{\sim}{\theta}_1, \ldots, \underset{\sim}{\theta}_N$. In the fixed-regressors model, a sample of N observations $\underset{\sim}{y}_1, \ldots, \underset{\sim}{y}_N$ of $\underset{\sim}{y}$ is thought of as corresponding to N fixed values, $\underset{\sim}{\theta}_1, \ldots, \underset{\sim}{\theta}_N$ of $\underset{\sim}{\theta}$, which are estimated jointly with the coefficients of the regression. From one point of view, there is an $n \times N$ matrix sample of size one from a peculiar conceptual universe that consists of all sets of N examinees with the same (unknown) values of $\underset{\sim}{\theta}_1, \ldots, \underset{\sim}{\theta}_N$ as in the sample that has been drawn.

In the terminology of Neyman and Scott (1948), the coefficients of the regression are structural parameters, while the latent trait values are incidental parameters whose number increases with N. Neyman and Scott showed that in a model containing such incidental parameters, the structural parameters may not be consistently estimated. In the linear case of the analogous multivariate regression model in which the independent variables are known quantities, the choice between a fixed-regressors model and a random-regressors model rests essentially on an application of the logic of the distinction. That is, the decision depends on whether the values of the independent variables are in fact randomly drawn with the experimental units or fixed by experimental control. (Some investigators seem unaware of the distinction and apply the statistics of the fixed-regressors model when they have random independent variables.) In the nonlinear case, the investigator will be likely to use the fixed-regressors model, even when the logic of the research design calls for random regressors, simply because the nonlinear random-regressors theory is difficult or impossible to work out. It is this latter consideration that seems to determine the choice in latent trait theory, where the regressor variables are defined by the principle of local independence. It seems unlikely, otherwise, that employing a fixed-regressors model in latent trait theory would ever be considered, since, rather obviously, a randomly drawn examinee brings his/her latent trait value(s) with him/her, so to speak.

Random-regressors and fixed-regressors treatments of latent trait theory each have their virtues and limitations. The next section examines some random-regressors models, while the section following it concerns fixed-regressors models.

Random-Regressors Theory

The first effective treatment of a latent trait model as a random-regressors model is possibly Lord's (1952) demonstration that if a set of binary items fits the normal ogive model and the latent trait is a normally distributed random variable, then the tetrachoric correlations of the items can be explained by the Spearman single common factor model. It is an open question whether the "heuristic" method, based on this result, of estimating the item parameters of the normal ogive model using a Spearman factor analysis of tetrachoric correlations has been or can be substantially improved upon, in terms of actual numerical results.

McDonald (1967a) pointed out that a (strictly) nonlinear latent trait model such as the normal ogive and latent distance models can be approximated as closely as desired by a polynomial series (a wide-sense linear model) on the

basis of harmonic analysis (Fourier analysis), a traditional device in the physical sciences for reducing a nonlinear model to a linear model. The coefficients of the polynomial series are so chosen that any finite segment of it is a least squares approximation to the desired strictly nonlinear model, weighted by the density function of the latent trait.

The main application of this method was to the normal ogive model, under the assumption that the latent trait is a random regressor with a normal distribution. In this case, McDonald (1967a) showed that if the normal ogive model is written as

$$\hat{y}_j = N(f_{j0} + f_{j1}\theta) \qquad [8]$$

where $N(\cdot)$ is the (unit) cumulative normal distribution function, then a weighted least-squares approximation to the model is given for any choice of $\underline{r}$ by

$$\phi_j^{(r)} = \sum_{p=0}^{r} b_{jp} h_p(\theta) \qquad [9]$$

where $h_p(\theta)$ is the normalized Hermite-Tchebycheff polynomial of degree $\underline{p}$, given by

$$h_p(\theta) = \frac{1}{\sqrt{p!}} \sum_{t=0}^{q} (-)^t \frac{\theta^{p-2t}}{2^t t! p-2t!} \qquad [10]$$

with $q = s$ if either $p = 2s$ or $p = 2s + 1$

$$b_{j0} = N\{f_{j0}/(1 + f_{j1}^2)^{1/2}\} \qquad [11]$$

and

$$b_{jp} = \frac{1}{\sqrt{p}} \left[\frac{f_{j1}}{(1 + f_{j1}^2)^{1/2}}\right]^p h_{p-1}\left[\frac{f_{j0}}{(1 + f_{j1}^2)^{1/2}}\right] n\{f_{j0}/(1 + f_{j1}^2)^{1/2}\} \qquad [12]$$

where $n(\cdot)$ is the (unit) normal density function. McDonald (1982a) has shown that the normal ogive is well approximated by the cubic polynomial obtained by terminating the series Equation 9 at $p = 3$. McDonald (1967a) proposed to fit the normal ogive model by estimating the first few coefficients of the series Equation 9 by nonlinear factor analysis, and solving sample analogues of equations 11 and 12 for f_{j0}, f_{j1}. Unpublished work by McDonald and Ahlawat (1974) showed that satisfactory estimates of f_{j0}, f_{j1} can be obtained from a (linear) Spearman factor analysis of the item covariance matrix to yield estimates of the quantities b_{j0}, b_{j1} (without obtaining coefficients of higher degree terms) and simultaneously solving the two equations 11 and

$$b_{j1} = \left\{ \frac{f_{j1}}{(1 + f_{j1}^2)^{1/2}} \right\} n \left\{ \frac{f_{j0}}{(1 + f_{j1}^2)^{1/2}} \right\} \qquad [13]$$

for f_{j0}, f_{j1} (substituting sample analogues). Thus, from this work a second "heuristic method" based on a Spearman factor analysis of the items is obtained.

Bock and Lieberman (1970) were able to obtain maximum likelihood estimates of the item parameters f_{j0}, f_{j1} in the normal ogive model, again assuming that the latent trait is a random regressor with a normal distribution. The method used a point approximation to the continuous distribution of the latent trait to yield a direct attack on the evaluation of the required integrals. The method was limited in application by the heavy computing demands in obtaining these approximations.

Christoffersson (1975) treated what may be recognized as a random-regressors multidimensional normal ogive model, using theory having elements in common with that of McDonald (1967a) and of Bock and Lieberman (1970). Christoffersson (1975) defined the model by introducing a set of n theoretical continuous variables that fit a multiple common factor model, and by supposing that the n observed binary item responses arise by dichotomization of these theoretical continuous variables at n threshold values to be estimated along with the parameters of the "underlying" common factor model. An expression is easily obtained for the proportion of examinees passing each item as a function of the model parameters. The proportion passing a pair of items can be approximated as closely as desired by taking sufficient terms of Pearson's tetrachoric series. (Christoffersson chose to stop at 10 terms.) Sample estimates of these proportions were then used to fit the model by generalized least squares. Muthén (1978) suggested the generalized least squares estimators in Christofferson's (1975) model can be obtained with less numerical cost by using sample tetrachoric correlations in place of sample proportions passing pairs of items.

McDonald (1980, 1982a, 1982b) has shown that the unidimensional normal ogive model can be fitted by ordinary least squares, as an application of a general purpose program for the analysis of covariance (moment) structures, using the harmonic analysis results of Equations 8 through 13. Instead of first fitting a Spearman model to the item covariances and then solving sample analogues of Equations 11 and 13 for the item parameter estimates, the covariance structure implied by Equation 9, with the coefficients given by Equations 11 and 12, is fit minimizing the least squares function with respect to the fundamental parameters f_{j0}, f_{j1}, $j = 1, \ldots, n$.

So far this section has consisted essentially of a review of known results. The remainder of this section will concern an extension of McDonald's (1967a) harmonic analysis of the unidimensional normal ogive to the multidimensional case and will relate this theory to Christoffersson's (1975) dichotomization of the common factor model. A multidimensional normal ogive model with a linear combination rule for the latent traits is defined as

$$\hat{y}_j = N\{f_{j0} + f_j'\theta\} = N\{f_{j0} + f_{j1}\theta_1 + \ldots + f_{jk}\theta_k\} , \qquad [14]$$

where, as before, $N(\cdot)$ is the cumulative normal distribution function. It is assumed that $\underset{\sim}{\theta}$ is a random regressor with a k-variate normal distribution, and a metric is chosen such that each component of $\underset{\sim}{\theta}$ has mean zero and variance unity. $\underset{\sim}{P}$ is written for the covariance (correlation) matrix of $\underset{\sim}{\theta}$ and

$$\underset{\sim}{F} = \begin{bmatrix} \underset{\sim}{f}_1' \\ \vdots \\ \underset{\sim}{f}_n' \end{bmatrix} \qquad [15]$$

where $\underset{\sim}{f}_j' = [f_{j1}, \ldots, f_{jk}]$. In some applications a pattern may be prescribed for $\underset{\sim}{F}$, with elements constrained to be zero (to define a "simple structure") or subject to other desired constraints, as in the counterpart factor-loading matrix in multiple factor analysis. In particular, by constraining the nonzero elements in each column to be equal (while choosing a pattern such that they are linearly independent), a multidimensional counterpart of the Rasch model is obtained. For each item score y_j, the item characteristic function $N(f_{j0} + \underset{\sim}{f}_j'\underset{\sim}{\theta})$ is constant on planes of dimension $k - 1$ orthogonal to the vector $\underset{\sim}{f}_j$. In the Appendix it is shown in detail that Equation 14 may be represented by the infinite polynomial series.

$$\hat{y}_j = \phi_j^{(\infty)}(\underset{\sim}{\theta}) = \sum_{p=0}^{\infty} b_{jp} h_p \left(\frac{1}{d_j} \underset{\sim}{f}_j'\underset{\sim}{\theta}\right) , \qquad [16]$$

where b_{j0}, b_{jp} are obtained by substituting

$$d_j = (\underset{\sim}{f}_j'\underset{\sim}{P}\underset{\sim}{f}_j)^{1/2} \qquad [17]$$

for f_{j1} in Equations 11 and 12, and $h_p(')$ is given by Equation 10 as before. The first $\underline{r}$ terms of the series Equation 16, which will be denoted by $\phi_j^{(r)}(\underset{\sim}{\theta})$, yield a polynomial of degree $\underline{r}$, which, like the multidimensional normal ogive Equation 14, is constant on planes of dimension $k - 1$ orthogonal to the vector $\underset{\sim}{f}_j$ and yields a weighted least squares approximation to it, as in the unidimensional case treated by McDonald (1967a). It further follows that the proportion of examinees passing an item is given by

$$\underset{\sim}{\pi}_j = \underset{\sim}{b}_{j0} \qquad [18]$$

and the proportion passing two items $\underline{j}$ and $\underline{k}$ is given by

$$\underset{\sim}{\pi}_{jk}^{(\infty)} = \sum_{p=0}^{\infty} b_{jp} b_{kp} \left(\frac{1}{d_j} \underset{\sim}{f}_j'\underset{\sim}{P}\underset{\sim}{f}_k \frac{1}{d_k}\right)^p , \qquad j \neq k = 1, \ldots, n . \qquad [19]$$

In practice the first $\underline{r}$ terms of Equation 19 may be substituted, denoted by $\pi_{jk}^{(r)}$, as a finite approximation to it. The model may be fit by least squares, minimizing the squared differences between sample proportions p_{jk} passing items $\underline{j}$ and $\underline{k}$, and $\pi_{jk}^{(r)}$ given by Equation 19, with respect to f_{j0}, $\underset{\sim}{f}_j$.

In the unidimensional case, Equation 19 reduces to the simple form

$$\pi_{jk}^{(\infty)} = \sum_{p=0}^{\infty} b_{jp}\, b_{kp}, \quad j \neq k = 1, \ldots, n , \qquad [20]$$

where b_{jp} is given by Equations 11 and 12, and the moment structure Equation 20 is then easily fitted by an application of McDonald's (1978, 1980) comprehensive model for the analysis of covariance structures, using a facility in the computer program for this work to evaluate b_{jp} and corresponding derivatives, as functions of f_{j0}, f_{j1}. In contrast, the moment-structure Equation 19 does not seem to yield an advantageous matrix expression and is comprehended by the comprehensive model only in the sense that the bivariate moments themselves are programmable functions of the model parameters. The simple linear approximation

$$\pi_{jk}^{(1)} = b_{j0}\, b_{k0} + \left[\frac{b_{j1}}{d_j}\right] \underset{\sim}{f}_j' \underset{\sim}{P} \underset{\sim}{f}_k \left[\frac{b_{k1}}{d_k}\right], \qquad j \neq k = 1, \ldots, n , \qquad [21]$$

may be written in matrix form as

$$[\pi_{jk}^{(1)}] = \underset{\sim}{b}_o \underset{\sim}{b}_o' + \underset{\sim}{D}^{-1}\, \underset{\sim}{D}_b \underset{\sim}{F} \underset{\sim}{P} \underset{\sim}{F}' \underset{\sim}{D}_b \underset{\sim}{D}^{-1} + \underset{\sim}{D}_u \qquad [22]$$

where

$$\underset{\sim}{b}_0' = [b_{10}, \ldots, b_{n0}]',$$
$$\underset{\sim}{D} = \text{Diag}\,\{d_1, \ldots, d_n\},$$
$$\underset{\sim}{D}_b = \text{Diag}\,\{b_{11}, \ldots, b_{n1}\}, \text{ and}$$
$$\underset{\sim}{D}_u = \text{Diag}\,\{b_{10}(1-b_{10}) - b_{11}^2, \ldots, b_{n0}(1-b_{no}) - b_{n1}^2\}.$$

Monte carlo studies of the unidimensional case have shown that the precision of estimation of the item parameters does not depend upon the number of terms retained in the polynomial approximation. On the other hand, the residuals $p_{jk} - \pi_{jk}^{(r)}$ tend, of course, to reduce in absolute values as $\underline{r}$ is increased, though convergence is very rapid, and terms beyond the cubic in general seem negligible. A simple and probably numerically efficient way to fit the model is to fit the linear approximation Equation 21 in the form Equation 22 using McDonald's general model for the analysis of covariance structures and then to use Equation 19 to some suitable order in order to evaluate the residuals more precisely.

However, at this point it should be noted that the theory just described is much closer to that of Christoffersson (1975) than might be obvious at first

sight. Indeed, the expression Equation 19 is identical with Equation 27 below, given by Christoffersson, except for a reparameterization of the model. Christoffersson defined a set of unobservable variables, $\underset{\sim}{v}$, that follow the multiple common factor model, say,

$$\underset{\sim}{v}_j = \underset{\sim}{\lambda}_j'\underset{\sim}{\theta} + \delta_j \qquad [23]$$

where

$$\underset{\sim}{\Lambda} = \begin{bmatrix} \underset{\sim}{\lambda}_1' \\ \vdots \\ \underset{\sim}{\lambda}_n \end{bmatrix} \qquad [24]$$

$\underset{\sim}{\Lambda}$ is a $n \times k$ matrix of common factor loadings,
$\underset{\sim}{\theta}$ is a vector of common factors, and
δ_j is the jth unique factor.

He then supposed that

$$\begin{aligned} y_j &= 1 \quad \text{if } v_j \geq t_j \\ &= 0 \quad \text{if } v_j < t_j \ . \end{aligned} \qquad [25]$$

It then follows that the proportion of examinees passing item j is given by

$$\pi_j = N(t_j) \qquad [26]$$

and the joint proportion of examinees passing items j and k is given by

$$\pi_{jk} = \sum_{p=0}^{\infty} \tau_p(t_j)\tau_p(t_k)(\underset{\sim}{\lambda}_j'P\underset{\sim}{\lambda}_k)^p \ , \qquad [27]$$

where $\tau_p(\cdot)$ is the tetrachoric function defined by

$$\tau_0(t_j) = N(t_j) \qquad [28]$$

$$\tau_p(t_j) = \frac{1}{\sqrt{p}}\, h_{p-1}(t_j)n(t_j) \ , \quad p = 1, 2, \ldots, \qquad [29]$$

and, as before, $h_p(\cdot)$ is the normalized Hermite-Tchebycheff polynomial Equation 10 and $n(\cdot)$ is the normal density function. Using Equations 28 and 29, the result (Equation 19) obtained by harmonic analysis of the model Equation 14 is rewritten as

$$\pi_{jk}^{(\infty)} = \sum_{p=0}^{\infty} \tau_p \left[\frac{f_{j0}}{(1+d_j^2)^{1/2}}\right] \tau_p \left[\frac{f_{k0}}{(1+d_k^2)^{1/2}}\right] \left[(1+d_j^2)^{-1/2} \underset{\sim}{f}_j' \underset{\sim}{\sim} P \underset{\sim}{f}_k (1+d_k^2)^{-1/2}\right] \qquad [30]$$

It is then immediately evident that the expressions Equations 27 and 30 are identical except for a choice of parameterization. Each is obtained from the other by writing

$$f_{j0} = \frac{t_j}{\sqrt{1-\underset{\sim}{\lambda}_j' \underset{\sim}{\sim} P \underset{\sim}{\lambda}_j}} \; ; \; \underset{\sim}{f}_j = \frac{\underset{\sim}{\lambda}_j}{\sqrt{(1-\underset{\sim}{\lambda}_j' \underset{\sim}{\sim} P \underset{\sim}{\lambda}_j}} \qquad [31]$$

or, conversely,

$$t_j = \frac{f_{j0}}{\sqrt{1+\underset{\sim}{f}_j' \underset{\sim}{\sim} P \underset{\sim}{f}_j}} \; ; \; \underset{\sim}{\lambda}_j = \frac{\underset{\sim}{f}_j}{\sqrt{1+\underset{\sim}{f}_j' \underset{\sim}{\sim} P \underset{\sim}{f}_j}} \; . \qquad [32]$$

To put this another way, Christoffersson's work implies writing the multidimensional normal ogive model as

$$\hat{y}_j = \frac{N\{t_j + \underset{\sim}{\lambda}_j' \underset{\sim}{\theta}\}}{\sqrt{1-\underset{\sim}{\lambda}_j' \underset{\sim}{\sim} P \underset{\sim}{\lambda}_j}} \} \qquad [33]$$

in place of

$$\hat{y}_j = N\{f_{j0} + \underset{\sim}{f}_j' \underset{\sim}{\theta}\} \qquad [34]$$

(It should be noted that in the unidimensional case, Bock and Lieberman, 1970, employ both these parameterizations.)

The equivalence of Christoffersson's (1975) tetrachoric series to McDonald's (1967a) harmonic analysis has both theoretical and practical consequences, apart from the possibility that a few research workers interested in IRT but not in factor analysis may have missed the point that Christoffersson's (1975) "factor analysis of dichotomized variables" is indeed the random-regressors multidimensional normal ogive model with linear combination rule. Purely in terms of theory it is pleasing to find that the same result, Equations 27 and 30, can be obtained either as the solution to the problem of evaluating a double integral connected with the normal distribution or as a solution to the problem of approximating a strictly nonlinear regression function by a wide-sense linear regression function. The main practical implication of the result is that it gives grounds for reducng numerical work by using what at first sight would appear to be unacceptably crude approximations, such as the simple first-term approximation in Equation 21. That is, fitting an approximation to the model may be considered instead of fitting the model itself, thereby estimating the parameters reasonably precisely, where there might not have been an expectation of obtaining reasonably precise estimates of the parameters using very crude ap-

proximations to the integrals required by the model itself.

It may prove to be the case that any of the random-regressors methods discussed in this section can be expected to give reasonably satisfactory estimates of the item parameters from large samples if the assumption of a univariate or multivariate normal distribution of the latent trait is not too grossly violated. All of them allow the computation and inspection of residual covariances. It may be argued (McDonald, 1981, 1982b) that on the basis of the magnitudes of the residual covariances, a direct judgment of the adequacy of fit of the model (including the crucial question of the dimensionality of the data) is of more importance than a Neyman-Pearson statistical decision to reject or not reject the model at the given sample size, since as a restrictive statistical hypothesis the model is never correct in applications to real data. The use of the "heuristic" methods that do not yield a test of significance can thus be rationalized, at least for large data sets for which such a test is not available in practice.

Certainly one limitation of the random-regressors treatment is that imposed by the normal distribution assumption. It seems to be regarded as a more dangerous assumption for item response models than in its ubiquitous applications to linear common factor analysis and other models for the structural analysis of multivariate data. Another general limitation, briefly described in the introduction, is that random-regressors theory for a wide range of item characteristic curves is difficult to develop, while the corresponding fixed-regressors theory is generally straightforward. Even the logistic model, which has a number of mathematical properties that cause it to be preferred to the numerically almost indistinguishable normal ogive model, does not lend itself so readily to treatment as a random-regressors model. Also, both Christoffersson's (1975) "factor-analytic" theory and the multidimensional theory based on McDonald's (1967a) harmonic analysis yield a random-regressors treatment for the multidimensional normal ogive model with a linear combination rule for the $\underline{k}$ latent traits, but it is not at all obvious how to extend this treatment to allow nonlinear combination rules, as in the model:

$$\hat{y}_j = N\{f_{j0} + f_{j1}\theta_1 + f_{j2}\theta_2 + f_{j3}\theta_1\theta_2\} \qquad [35]$$

or

$$\hat{y}_j = N\{f_{j0} + f_{j1}\theta_1 + f_{j2}\theta_1^2\} . \qquad [36]$$

To see how such models might be dealt with the next section will be concerned with fixed-regressors theory.

Fixed-Regressors Theory

In this section no attempt will be made to review previous work. It must suffice to note that direct attacks on the problem of jointly estimating item parameters and examinee "abilities" in the normal ogive, logistic, and Rasch models by maximum likelihood have been made by Kolakowski and Bock (1970), Wingersky and Lord (1973), and Wright and his associates (Wright & Stone, 1979), at

least. Although there is some evidence from monte carlo studies that these methods yield satisfactory estimates of the item parameters, it seems broadly true that they have not been made to yield a satisfactory test of fit, since the likelihood function has not been made to yield a loss function, measuring poorness of fit, and no statistics are computed to test for departure in the sample from obedience to the principle of local independence, given the hypothesized (unit) dimensionality. (Wright and Stone have computed fit statistics that are functions of the residuals of each item. These quantities do not bear any obvious relationship to the principle of local independence.) Consider next fixed-regressors theory for unidimensional and multidimensional latent trait models based directly on the nonlinear common factor model Equation 1, with the weak principle of local independence. The general model Equation 1 is rewritten as a fixed-regressors model,

$$y_{ji} = \phi_j(\theta_{1i}, \ldots, \theta_{ki} ; \beta_{j1}, \ldots, \beta_{js}) + e_{ji},$$

$$j = 1, \ldots, n ; i = 1, \ldots, N , \qquad [37]$$

where $\beta_{j1}, \ldots, \beta_{js}$ is a set of parameters describing the jth regression function. The residual covariances q_{jk} are defined by

$$q_{jk} = \frac{1}{N} \sum_i e_{ji} e_{ki} , \qquad [38]$$

and the (weak) principle of local independence is expressed as

$$\mathcal{E}\{q_{jk}\} = u_j^2 \quad , j = k ,$$

$$= 0 \quad , j \neq k . \qquad [39]$$

The $n \times n$ matrix of sample residual covariances is written $\underset{\sim}{Q} = [q_{jk}]$, and $U^2 = \text{Diag}\left\{u_1^2, \ldots, u_n^2\right\}$, the expected value of $\underset{\sim}{Q}$ under the hypothesis.

McDonald (1979, in press) shows that any model of the type Equation 37 may be fit by choosing values of $\theta_{1i}, \ldots, \theta_{ki}, \beta_{j1}, \ldots, \beta_{js}, u_1^2, \ldots, u_n^2$, to make $\underset{\sim}{Q}$ as close as possible to a diagonal matrix in one of two senses. With respect to these parameters, either the loss function

$$\omega = \tfrac{1}{2}\text{Tr}\{(\underset{\sim}{Q} - \underset{\sim}{U}^2)^2\} , \qquad [40]$$

or the loss function

$$\ell = -\tfrac{1}{2} \log\left|(\text{Diag } \underset{\sim}{Q})^{-\frac{1}{2}} \underset{\sim}{Q} (\text{Diag } \underset{\sim}{Q})^{-\frac{1}{2}}\right| . \qquad [41]$$

may be minimized. The first of these may be recognized as Harman's well-known MINRES loss function in linear common factor analysis, by the use of which there is an endeavor to minimize the squares of the residual covariances of distinct variables. The second corresponds to the maximum determinant method in linear

factor analysis (Howe, 1955). By its use there is an endeavor to maximize the determinant of the correlation matrix of the residuals, thus making it as close as possible to an identity matrix. If the residuals can be assumed to have a multivariate normal distribution, which is certainly not the case for binary items, the second loss function also yields maximum likelihood ratio estimators (McDonald, 1979).

In principle, any prescribed model of the type Equation 37 can be fitted to binary items or test scores by minimizing either of the loss functions Equations 40 and 41; the goodness of fit of the model may be judged on the basis of the magnitudes of the loss functions. An inspection of the arrangement of magnitudes of the residual covariances in the residual covariance matrix may indicate the source of any disagreement between the model and the data and may suggest a revised model. To fit any model whatsoever of type Equation 37, it is a straightforward matter to program expressions for the loss function and its derivatives with respect to the parameters of the model and to apply a minimization algorithm, such as the method of conjugate directions, from a suitable starting point. It proves to be necessary to alternate between minimization of one of the loss functions Equation 40 or 41 with respect to the examinee parameters θ_{ji}, and minimization of the usual mean-square residual

$$\tau = \mathrm{Tr}\{\underset{\sim}{Q}\} \qquad [42]$$

with respect to the parameters of the regression on the current θ_{ji} values. It may be shown (see McDonald, in prep.) that the minimum point of τ with respect to the parameters of the regression is the same as the minimum of Equation 40 or 41 with respect to those parameters. Minimization of Equation 40 or 41 with respect to the parameters of the regression yields poor estimates in general. Curiously, the better the fit to the model, the poorer such estimates would be, and alternating the loss functions would be better.

Thus far, the only case of the model Equation 37 that has been extensively investigated is a polynomial model with pairwise interaction terms (Etezadi-Amoli & McDonald, in press; McDonald, 1979), i.e., a model of the type

$$y_{ji} = \sum_{p=1}^{k} \sum_{q=1}^{r_p} \beta_{jpq} \theta_p^q + \sum\sum_{p \neq p'} \gamma_{jpp'} \theta_p \theta_{p'} + e_{ji} , \qquad [43]$$

a wide-sense linear model which includes as a special case the simple polynomial model

$$y_{ji} = \sum_{q=1}^{r} \beta_{jq} \theta^q + e_{ji} . \qquad [44]$$

However, unlike random-regressors theory, the fixed-regressors theory just described should be just as easily applied to strictly nonlinear models, for example, to the multidimensional normal ogive model Equation 14 rewritten as

$$y_{ji} = N(f_{j0} + \underset{\sim}{f}_j' \underset{\sim}{\theta}_i) + e_{ji} , \qquad [45]$$

where, if the observed variables are binary item responses, the conditional distribution of y_{ji} given $\underset{\sim}{\theta}_i$ is binomial, with heteroscedastic variances given by

$$\mathcal{E}\{e_{ji}^2 \mid \underset{\sim}{\theta}_i\} = N(f_{j0} + \underset{\sim}{f}_j\underset{\sim}{\theta}_i) \quad [1 - N(f_{j0} + \underset{\sim}{f}_j'\underset{\sim}{\theta}_i)] \ , \qquad [46]$$

and certainly not normal. Again, as in the random-regressors form of this model, a pattern for the parameters $\underset{\sim}{f}_j$ corresponding to simple structure may be prescribed or they may be constrained to be equal in columns of the matrix

$$\begin{bmatrix} \underset{\sim}{f}_1' \\ \vdots \\ \underset{\sim}{f}_n' \end{bmatrix} \qquad [47]$$

to yield a multidimensional counterpart of the Rasch model. In contrast to the random-regressors form of the model, any convenient and appropriate nonlinear function for $N(\cdot)$ may be substituted--the logistic function, for example--without creating any mathematical problems for the implementation of the theory. Perhaps more interestingly, the fixed-regressors version of the model allows escape from the restrictions of the linear combination rule. The model $N(f_{j0} + \underset{\sim}{f}_j'\underset{\sim}{\theta}_i)$ is a transformation, suitable for binary items, of the linear common factor model; and, as noted earlier, the regression functions are constant on planes orthogonal to $\underset{\sim}{f}_j$. Whereas the application of random-regressors theory to models such as Equation 35 and 36 seems to yield intractable mathematical problems, the fixed-regressors versions of such models

$$y_{ji} = N(f_{j0} + f_{j1}\theta_{1i} + f_{j2}\theta_{2i} + f_{j3}\theta_{1i}\theta_{2i}) + e_{ji} \qquad [48]$$

and

$$y_{ji} = N(f_{j0} + f_{j1}\theta_{1i} + f_{j2}\theta_{1i}^2) \qquad [49]$$

should be just as easy to fit as the model Equation 14, so the limitations of the linear combination rule are easily escaped.

A great deal of numerical work is needed on models arising out of the fixed-regressors treatment of Equations 1 and 2 by the weak principle of local independence before there can be confidence that the approach is generally practical. The virtues of this approach consist in its generality, its freedom from distribution assumptions, and the fact that it directly embodies the weak principle of local independence and thus tests the agreement of the chosen model, with prescribed dimensionality, with the data. An obvious disadvantage is that even with the use of an efficient minimization algorithm, such as the method of conjugate directions, limitations of sample size, which may be serious given the poor quality of information in binary data, are immediately apparent, whereas random-regressors methods are not limited by sample size.

A Numerical Example

The fixed-regressors methods described above have not yet been programmed for regression curves, unidimensional or multidimensional, that are suitable for binary responses. A program for the unidimensional normal ogive model based on McDonald's (1967a) harmonic analysis has been written by Fraser, who has recently extended this work to cover the multidimensional model Equation 14. A detailed account of the method used in these programs and of their numerical behavior will be given elsewhere (McDonald & Fraser, in prep.).

Here, a single numerical illustration of the multidimensional version of the method will be given to illustrate the general claim made above--that even the linear approximation can yield reasonable estimates because of the weighted least squares property of the polynomial series approximation to the normal ogive.

Let p_j be the proportion of examinees passing item j, and p_{jk} be the proportion passing items j and k, in a sample of size N. In program NOHARM (Normal Ogive Harmonic Analysis Robust Method) the threshold or position parameter is estimated in closed form by solving the sample analogue of Equation 26, i.e., by obtaining

$$\hat{t}_j = N^{-1}(p_j) \ , \qquad [50]$$

and the parameters $\underset{\sim}{f}'_j$ are obtained by ordinary least squares, minimizing

$$\phi = \sum_{j \neq k} (p_{jk} - \pi_{jk}^{(r)})^2 \ , \qquad [51]$$

where $\pi_{jk}^{(r)}$ is the r-term approximation to $\pi_{jk}^{(\infty)}$ in Equation 19, by a quasi-Newton algorithm. The combination of the random-regressors model and the weak principle of local independence with the use of ordinary least squares makes it possible to analyze quite large numbers of items with unlimited sample size. [With commonly available computer configurations, it should be possible to fit an equivalent of the Rasch model to 200 items, and the 2-parameter model (or the 3-parameter model with supplied pseudo-guessing parameters) to more than 100 items, with an unlimited number of examinees.

Data from sections of the LSAT have been used by a number of writers to illustrate IRT. In particular, LSAT7 has been treated by Christoffersson (1975) as a two-dimensional case of the normal ogive model. The sample raw product-moment matrix from LSAT7, with N = 1,000, is given in Table 1. Program NOHARM (McDonald & Fraser, in prep.) was used to fit model Equation 14 to the data, first using the linear approximation, then using the cubic approximation to the harmonic series. In both cases, orthogonal latent traits were prescribed, and an unrestricted 5 × 2 matrix of the parameters $\underset{\sim}{f}'_j$, j = 1, ..., 5, in Equation 14 was fitted and then rotated by varimax for comparison with Christoffersson's result. Table 2 gives the estimates from the linear and cubic approximation,

Table 1
Sample Raw Product-Moment Matrix for Five Items from LSAT-7

Item	1	2	3	4	5
1	.828				
2	.567	.658			
3	.664	.560	.772		
4	.532	.428	.501	.606	
5	.718	.567	.672	.526	.843

after varimax rotation, in the parameterization Equation 14, while Table 3 gives the corresponding results after reparameterization by Equation 32, together with Christoffersson's (1975) generalized least squares estimates.

Table 2
Parameter Estimates from NOHARM

	Linear Approximation			Cubic Approximation		
Item	f_{j0}	f_{j1}	f_{j2}	f_{j0}	f_{j1}	f_{j2}
1	2.101	1.966	.253	1.975	1.819	.213
2	.491	.231	.636	.483	.231	.594
3	1.361	.493	1.446	1.270	.420	1.314
4	.295	.345	.289	.295	.341	.295
5	1.099	.373	.227	1.086	.331	.234

Table 3
Reparameterized Estimates

	Linear			Cubic			Christofferson		
Item	t_j	λ_{j1}	λ_{j2}	t_j	λ_{j1}	λ_{j2}	t_j	λ_{j1}	λ_{j2}
1	-.946	.886	.114	-.946	.872	.102	-.950	.796	.146
2	-.399	.191	.527	-.407	.195	.501	-.406	.197	.475
3	-.746	.270	.792	-.745	.246	.771	-.747	.212	.826
4	-.260	.315	.264	-.269	.311	.269	-.270	.325	.258
5	-.998	.342	.208	-1.007	.307	.213	-1.007	.310	.225

Table 4 gives the residual covariance matrices for the linear approximation and the cubic approximation. In this case, the results from the latter can hardly be claimed to be an improvement on those from the former, and both are in reasonable agreement with Christoffersson's results based on a 10-term series. The example seems to support the claim made earlier that the normal ogive model is remarkably well fitted by fitting the approximating linear model.

Table 4
Residual Covariance Matrices for Linear Approximation (Lower Triangle) and Cubic Approximation

Item	1	2	3	4	5
1		-.0006	.0002	.0002	.0002
2	-.0007		-.0001	-.0012	.0031
3	.0006	.0000		.0010	-.0020
4	.0001	-.0012	.0010		-.0007
5	.0000	.0031	-.0025	-.0001	

Conclusions

It has been suggested that there are a number of advantages to be obtained from treating item response theory as a special case of nonlinear common factor analysis. Nonlinear common factor analysis supplies a general yet rigorous definition of the dimensionality of a set of tests or of a set of dichotomous or polychotomous items, as the number of common factors or latent traits required in the model. This definition replaces ill-defined notions of homogeneity and internal consistency which have persisted in the literature, without clear explication, as near synonyms for unidimensionality (see McDonald, 1981). Nonlinear common factor analysis yields a natural embodiment of the weak principle of local independence in appropriate loss functions that can be used to fit item response models and to assess their adequacy as descriptions of the data. The recognition of latent traits as synonymous with common factors gives appropriate guidance for interpretation, as well as for estimating and testing the model, in both unidimensional and multidimensional versions. Note, for example, that Christoffersson's (1975) factor-analytic parameterization of the multidimensional normal ogive model, Equation 33, suggests the use of factor analytic standards for "salient" factor loadings in the identification of latent traits.

Within limitations set by the assumption of a normal distribution of latent traits, random-regressors theory enables the fitting of item parameters in the unidimensional and multidimensional normal ogive model with unlimited sample sizes and quite large item sets with very satisfactory numerical efficiency. There is usually little interest in the values of the latent traits of the examinees in the calibration sample, and if there is interest, these can easily be estimated in a second stage after the item parameters have been estimated. Random-regressors treatments do, then, have much to recommend them.

Fixed-regressors theory based on McDonald's (1979) treatment of common factor analysis is, in principle, applicable to any prescribed unidimensional or multidimensional model in item response theory, though so far it has been applied only to the rather inappropriate polynomial model. The advantages of this approach are (1) that it is not limited to the use of a linear combination rule for multiple latent traits and (2) that it easily permits the introduction of interaction terms. Applications of the fixed-regressors treatment are subject

to limitations of sample size, which may prove severe in the case of binary data, where large sample sizes are generally desirable.

Another contribution of the factor analytic perspective on item response theory, not discussed in this paper, consists in the treatment of the invariance of item parameters across populations as closely related to the traditional treatments of factorial invariance (see McDonald, 1982b). Perhaps enough has been suggested to persuade researchers in the field of item response theory to recognize at least the theoretical value of seeing it as part of a more general psychometric unity.

REFERENCES

Bock, R. D., & Lieberman, M. Fitting a response model for n dichotomously scored items. Psychometrika, 1970, 35, 179-197.

Christoffersson, A. Factor analysis of dichotomized variables. Psychometrika, 1975, 40, 5-22.

Etezadi-Amoli, J., & McDonald, R. P. A second generation nonlinear factor analysis. Psychometrika, in press.

Howe, W. G. Some contributions to factor analysis (Report No. ONRL 1919). Oak Ridge TN: Oak Ridge National Laboratory, 1955.

Kolakowski, D., & Bock, R. D. A Fortran-IV program for maximum likelihood item analysis and test scoring: Normal ogive model (Research Memorandum No. 12). Chicago: University of Chicago, Department of Education, Statistical Laboratory, 1970.

Lazarsfeld, P. F., & Henry, N. W. Latent structure analysis. Boston: Houghton-Mifflin, 1968.

Lord, F. M. A theory of test scores. Psychometric Monographs, 1952, No. 7.

Lord, F. M., & Novick, M. R. Statistical theories of mental test scores. Reading MA: Addison-Wesley, 1968.

McDonald, R. P. A note of the derivation of the general latent class model. Psychometrika, 1962, 27, 203-206. (a)

McDonald, R. P. A general approach to nonlinear factor analysis. Psychometrika, 1962, 27, 397-415.

McDonald, R. P. Nonlinear factor analysis. Psychometric Monographs, 1967a, No. 15.

McDonald, R. P. Numerical methods for polynomial models in nonlinear factor analysis. Psychometrika, 1967, 32, 77-112.

McDonald, R. P. Factor interaction in nonlinear factor analysis. British Journal of Mathematical and Statistical Psychology, 1967, 20, 209-215.

McDonald, R. P. PROTEAN--A CDC computer program for nonlinear factor analysis (Research Memorandum RM-67-26). Princeton NJ: Educational Testing Service, 1967. (d)

McDonald, R. P. A simple comprehensive model for the analysis of covariance structures. British Journal of Mathematical and Statistical Psychology, 1978, 31, 59-72.

McDonald, R. P. The structural analysis of multivariate data: A sketch of a general theory. Multivariate Behavioral Research, 1979, 14, 21-38.

McDonald, R. P. A simple comprehensive model for the analysis of covariance structures: Some remarks on applications. British Journal of Mathematical and Statistical Psychology, 1980, 33, 161-183.

McDonald, R. P. The dimensionality of tests and items. British Journal of Mathematical and Statistical Psychology, 1981, 34, 100-117.

McDonald, R. P. Some alternative approaches to the improvement of measurement in education and psychology: Fitting latent trait models. In D. Spearritt (Ed.), The improvement of measurement in education and psychology. Australian Council for Educational Research, 1982. (a)

McDonald, R. P. Linear versus nonlinear models in latent trait theory. Applied Psychological Measurement, 1982, 6, 379-396. (b)

McDonald, R. P. Exploratory and confirmatory nonlinear factor analysis. In H. Wainer (ed.), Festschrift for F. M. Lord. Hillsdale NJ: Erlbaum, in press.

McDonald, R. P., & Ahlawat, K. S. Difficulty factors in binary data. British Journal of Mathematical and Statistical Psychology, 1974, 27, 82-99.

Muthén, B. Contributions to factor analysis of dichotomous variables. Psychometrika, 1978, 43, 551-560.

Neyman, J., & Scott, L. Consistent estimates based on partially consistent observations. Econometrika, 1948, 16, 1-32.

Wingersky, S., & Lord, F. M. A computer program for estimating examinee ability and item characteristic curve parameters when there are omitted responses (Research Memorandum RM-73-2). Princeton NJ: Educational Testing Service, 1973.

Wright, B. D., & Stone, M. H. Best test design. Chicago: Mesa Press, 1979.

APPENDIX

To obtain Equations 16 and 19 from Equation 14, the procedure is as follows: A transformation of $\underset{\sim}{\theta}$ into $\underset{\sim}{\theta}^*$ is sought, such that $\hat{y}_j$ is a function of one component of $\underset{\sim}{\theta}^*$, say the first, only. Define $\underset{\sim}{T}_j$ by

$$\underset{\sim}{T}_j = \frac{1}{\delta_j} \begin{bmatrix} \underset{\sim}{f}'_j \\ \underset{\sim}{f}'_{2j} \\ \vdots \\ \underset{\sim}{f}'_{kj} \end{bmatrix} \qquad [A1]$$

where

$$\delta_j = \underset{\sim}{f}'_j \underset{\sim}{f}_j \qquad [A2]$$

and $\underset{\sim}{f}_{\ell j}$ such that

$$\underset{\sim}{f}'_{\ell j} \underset{\sim}{f}_j = 0 \ , \qquad [A3]$$

$$\underset{\sim}{f}'_{\ell j} \underset{\sim}{f}_{mj} = 0 \ , \ \ell \neq m \ , \qquad [A4]$$

and

$$\underset{\sim}{f}'_{\ell j} \underset{\sim}{f}_{\ell j} = \delta_j^2 \ , \ \ell = 2, \ldots, k \ . \qquad [A5]$$

(That is, the rows of $\underset{\sim}{T}_j$ are completed arbitrarily after the first to yield an orthogonal matrix.)

Define

$$\underset{\sim}{\theta}^* = \frac{\delta_j}{d_j} \underset{\sim}{T}_j \underset{\sim}{\theta} \qquad [A6]$$

where

$$d_j^2 = \underset{\sim}{f}'_j \underset{\sim\sim}{P} \underset{\sim}{f}_j \qquad [A7]$$

so that, in particular,

$$\theta_1^* = \frac{1}{d_j} \underset{\sim}{f}_j' \underset{\sim}{\theta} \ , \qquad [A8]$$

with

$$\mathcal{E}\{\theta_1^*\} = 0 \ ; \ \mathcal{E}\{\theta_1^{*2}\} = 1 \qquad [A9]$$

(since $\mathcal{E}\left\{\underset{\sim}{f}_j'\underset{\sim}{\theta}\underset{\sim}{\theta}'\underset{\sim}{f}_j\right\} = \underset{\sim}{f}_j'\underset{\sim}{P}\underset{\sim}{f}_j$).

The converse transformation to Equation A6 is

$$\underset{\sim}{\theta} = \frac{d_j}{\delta_j}\,\underset{\sim}{T}_j'\underset{\sim}{\theta}^* \,, \qquad [A10]$$

since $\underset{\sim}{T}_j$ is orthogonal. Then

$$N\,(f_{j0} + \underset{\sim}{f}_j'\underset{\sim}{\theta}) = N(f_{j0} + \frac{d_j}{\delta_j}\,\underset{\sim}{f}_j'\underset{\sim}{T}_j'\underset{\sim}{\theta}^*) \;; \qquad [A11]$$

and since, by construction of $\underset{\sim}{T}_j$,

$$\underset{\sim}{f}_j'\underset{\sim}{T}_j' = [\delta_j \; 0 \; 0 \; \ldots \; 0] \qquad [A12]$$

gives, further,

$$N\,(f_{j0} + \underset{\sim}{f}_j'\underset{\sim}{\theta}) = N\,(f_{j0} + d_j\theta_1^*) \,. \qquad [A13]$$

Since $N(f_{j0} + d_j\theta_1^*)$ is a function of one variable only, by the properties of the unidimensional case, it may be represented by the infinite polynomial series

$$\phi_j^{(\infty)}\,(\theta_1^*) = \sum_{p=0}^{\infty} b_{jp}h_p(\theta_1^*) \qquad [A14]$$

where b_{j0}, b_{jp} are obtained by substituting d_j given by Equation A7 for f_{j1} in Equations 11 and 12. Equation 16 then follows.

Further, according to Lancaster (1958), if (x, y) are bivariate normally distributed with means zero, variances unity, and correlation ρ, then

$$\mathcal{E}\{h_p(x)h_q(y)\} = \rho^p \,, \; p = q$$

$$= 0 \,, \quad p \neq q \,, \qquad [A15]$$

where, as before, $h_p(\cdot)$ is the normalized Hermite-Tchebycheff polynomial given by Equation 10. It follows that

$$\pi_j = b_{j0} \qquad [A16]$$

and

$$\pi_{jk}^{(\infty)} = P\{y_j = 1, \; y_k = 1\} \,,$$

$$= \mathcal{E}\{N(f_{j0} + \underset{\sim}{f}_{j}'\theta)\ N(f_{k0} + \underset{\sim}{f}_{k}'\theta)\},$$

$$= \sum_{p=0}^{\infty} b_{jp} b_{kp} \left(\frac{1}{d_j} \underset{\sim}{f}_{j}' \underset{\sim\sim}{P} \underset{\sim}{f}_{k} \frac{1}{d_k}\right)^{P}, \qquad [A17]$$

as stated in Equation 19.